DOCUMENT RESUME

ED 365 153 FL 021 771

AUTHOR Nair-Venugopal, Shanta
TITLE Continuous Assessment in the Oral Communication Class: Teacher Constructed Test.
PUB DATE 91
NOTE 15p.; In: Sarinee, Anivan, Ed. Current Developments in Language Testing. Anthology Series 25. Paper presented at the Regional Language Centre Seminar on Language Testing and Language Programme Evaluation (April 9-12, 1990); see FL 021 757.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Second Language Programs; Communicative Competence (Languages); *English (Second Language); Evaluation Criteria; Foreign Countries; Higher Education; Language Teachers; *Language Tests; Oral Language; Rating Scales; Second Languages; *Teacher Developed Materials; Teacher Role; Test Anxiety; *Test Construction; Testing; Test Use; *Verbal Tests
IDENTIFIERS Malaysia; National University of Malaysia

ABSTRACT

The oral communication course for English majors at the National University of Malaysia includes testing designed by faculty and coordinated with the curriculum. This practice is based on the ideas that a teacher who has been actively involved in curriculum design is in a good position to design a test for that curriculum, and that teacher-made tests have a beneficial backwash effect on student learning. The course features two levels of instruction, each taught over two consecutive semesters. Final tests for both levels sample global communicative ability. Because the approach is communicative, the examinations are series of tests administered throughout the semester, allowing for continuous feedback to aid instruction. At level 1, the tests focus on three speaking tasks: extended, impromptu speech; group discussion; and an end-of-semester project. The tasks test three modes of speech: talking about oneself, others, experiences; narrating and describing events; and expressing and justifying opinions. At level 2, tests focus on group discussion, public speaking, debating, and an end-of-semester project. Rating scales have been constructed for all tests based on the types of communicative ability required. Continuous testing has reduced test anxiety. Test development is ongoing. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Weng

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 365 153

CONTINUOUS ASSESSMENT IN THE ORAL COMMUNICATION CLASS: TEACHER CONSTRUCTED TEST

Shanta Nair - Venugopal

INTRODUCTION

The Teacher As Tester

There is evidence that not only are teachers good judges of behaviour, they are also reliable judges of test performances. (Callaway, D R 1980). However it would be quite naive and perhaps even imprudent to suggest then, that all teachers will also by extension make naturally good testers given Spolsky's (1975) rhetoric on whether testing is art or science. Nevertheless, it can be assumed that a teacher who has been actively involved in course design or better still in the privileged position of 'negotiating' the curriculum, with her students would at least have a blueprint of sorts as a starting point for the construction of tests for that course. This could be further enhanced if the process is subjected to friendly criticism at the very least by other members of staff in relation to the objectives of the course or curriculum as a whole. The teacher is then in the informed and educated position of being able to translate the objectives of the course into tests construction by linking the specific objectives of the course with the task specifications identified. The test would then be underpinned by at least a view of language learning even if not a full fledged theory, in a clear case of doing the best that can be done. The analogy is best supplied by Skehan (1988) who summarized the current state of the art on (communicative) testing.

"...Since ... definitive theories do not exist, testers have to do the best they can with such theories as are available."

The contention therefore is that the teacher who has had some responsibility for course design and implementation is in many ways pre-eminently qualified to construct tests for the course particularly if it is backed by experience and shared knowledge in the field. Since the target group is known at first hand, needs can be fairly accurately specified on the basis of introspection and experience. The backwash effect of teacher-made tests on teaching can only be beneficial. As the teacher in this case is also responsible for course content (and like all other teachers across the board has the best interests of her students

FL021771

BEST COPY AVAILABLE

at heart), she will certainly teach what is to be tested, test what is taught and 'bias for best' in the use of test procedures and situations. The only possible danger lurking in this happy land is the possibility of a teacher who willy-nilly teaches the test as well and thereby nullifies its value as a measuring instrument.

BACKGROUND

The Target Group

At the English Department of the National University of Malaysia (UKM), students in the second year of the B A in English Studies program are required to take both levels 1 and 2 of an oral communication course that straddles two semesters or one academic session. These students are viewed as potential candidates for the B A in English Studies degree and there is a tremendous responsibility (equally shared by the writing and reading courses) to improve their language ability to make them "respectable" (Nair-Venugopal, S. 1988) candidates for the program. This may be seen as the perceived and immediate need. The projected or future need is seen as a high level of language ability that also makes for good language modelling as there is evidence that many of these students upon graduation enroll for a diploma in Education and become English language teachers. The mature students in the course are invariably teachers too. The responsibility is even more awesome given the language situation in the country which while overtly ESL also manifests many hybrids of the ESL/EFL situation, notwithstanding government efforts at promoting English as an important second language. These students (except those who are exempted on the basis of a placement test and have earned credits equivalent to the course) are also subject to a one year fairly intensive preparatory proficiency program (twelve hours per week). The emphasis in this course is on an integrated teaching of the four language skills. These students have also had a minimum of eleven years of instruction in English as a subject in school. There is also invariably the case of the mature student who has probably had 'more' English instruction, having been subject chronologically to a different system of education in the country's history.

Course Objectives

The oral communication course comprises two levels- each level taught over two semesters consecutively. The general aim of level I is to provide a

language learning environment for the acquisition of advanced oral skills and that of level II to augment and improve upon the skills acquired in level I, thus providing a learning continuum for the acquisition of advanced oral skills. At this juncture it must be pointed out that in the integrated program of the first year there is an oral fluency component. In other words the students in the second year have already been thrown into the 'deep end' as it were and the assumption is that upon entry to Level I they have more than banal or survival skills in oral communication. The reality is that students in spite of the first year of fairly intensive instruction and exposure enter the second year with varying levels of abilities. The task at hand for the second year oral skills programme is quite clear; raise levels of individual oral ability, bridge varying levels of individual abilities and yet help students to develop at their own pace. Hence the need to see the language class as a language acquisition environment bearing in mind that contact and exposure with the language outside the class is not optimal. The main objective in Level I is to achieve a high level of oral fluency in the language with an accompanying level of confidence and intelligibility, the latter being viewed with some urgency since native vernaculars are increasingly used for social communication outside the classroom and Bahasa Malaysia remains the language of instruction for courses in all other disciplines. The main objective of Level II is to achieve a high level of oral language ability. Both these objectives are further broken down into specific objectives for both levels. The tests are pegged against these objectives.

The specific objectives of Level I of the course are as follows:

1 attain high levels of intelligibility in speech

2 comprehend standard varieties of the spoken language without difficulty

3 interact and converse freely among themselves and other speakers of the language

4 convey information,narrate and describe; express and justify opinions.

These objectives are realized through an eclectic methodology using a variety of instructional devices, classroom procedures and multimedia materials.

The second objective is realized largely through practice in the language laboratory and it is not tested ie. elicited for as a skill domain in the tests that have been developed for the course. While it is generally accepted that listening comprehension as a skill is not easy to teach, it is even more elusive to test. According to Brown, G. and Yule, G. (1983)

> "...a listener's task performance may be unreliable for a number of reasons... we have only a very limited understanding of how we could determine what it is that listening comprehension entails. Given these two observations, it would seem that the assessment of listening comprehension is an extremely complex undertaking".

Having said that, why then has listening comprehension been included as a desirable objective on the course? As the view of language underlying the course is that of communication, no course that purports to teach oral communication (which view of language surely sees listening as a reciprocal skill) can justifiably not pay attention to teaching it at least. Objective 3 is specifically tested as speech interaction in the form of group discussions and 4 as extended "impromptu" speech in 3 modes. 1 is rated as a variable of performance for both these test types. 4 is also subsumed as 'enabling' skills in the group discussion test.

Objectives for level 2 are as follows:

1 not only comprehend all standard varieties of the language but also make themselves understood to other speakers of the language without difficulty.

2 participate in discussions on topics of a wide range of general interest without hesitation or effort

3 speak before audiences confidently (as in public speaking/platform activities)

4 convey information, persuade others and express themselves effectively as users of the language (as in debates and forums)

These objectives are achieved through the use of a selection of instructional devices, classroom procedures and modes such as simulations, small group discussions, debates and public speaking.

Objective 2 is tested using the group discussion test. 3 and 4 to borrow Tarone's notion (1982/83) of a "continuum of interlanguage styles" are to be seen as examples of "careful styles" and are tested as formal modes of speaking and debates. Objective 4 is also elicited as performance variables in the group discussion test. The second part of 1 ie. intelligibility/comprehensibility operates as an important variable in assessing the performance of all these tests. The final tests for both levels sample global communicative ability in the rehearsed speech genre which is an oral newsmagazine presentation on tape for the first

BEST COPY AVAILABLE

level and a videotaped presentation for the second level of either one of two platform activities or a chat show. Both are take-home, end-of-semester projects.

THE TESTS

Some Considerations

"In constructing tests, it is essential to have a defined curriculum or a set body of knowledge from which testers determine what to test (Shohamy, E 1988)".

To echo Charles Alderson (1983) the most important question to be asked of any test is, "What is it measuring?" which "can be determined by a variety of means including face inspection". Needless to say there are two other questions that merit equal consideration. One is, how is it measured and perhaps more crucially why? With reference to these tests, the question "for whom" ie. the target group has already been answered. As for purpose, each test type is seen as having a specified purpose that corresponds to an ability in an oral skill domain that has been delineated in the course objectives. Task specifications are prescribed by the oral skills domains. Therefore each test would sample different behaviour or skills in the form of different speech modes and the task specifications will vary from test type to test type. However all tests will test for both linguistic and communicative ability.

> "It is difficult to totally separate the two criteria, as the linguistic quality of an utterance can influence comprehensibility the basic communicative criterion. Further, while a major goal of most college or secondary language programs is communicative ability in the target language, there is justifiable concern with linguistic correctness because ...we are not just attempting to teach survival communications..., we are also trying to teach literacy in another language". Bartz W H (1979)

It is quite clear that as the view of the language underlying the teaching is communicative and the view of language learning, that of acquisition, achievement tests administered both mid-way and at the end of each semester will not allow the teacher to obtain feedback on acquired ability which could be used for diagnostic purposes as well (particularly at entry from the first level to the second), nor allow for a 'profiling' of performance. Hence the need for and

the development of a continuous 'battery' of tests, spaced out in relation to their ordering on the course and as spelt out by the course objectives. These have been conceptualized as oral skills domains and rated accordingly.

> "...Advances in the state of the art of achievement testing are directly related to advances in the concept of skills domains on which student achievement is assessed". Shoemaker (cited by Swain M. 1980)

The tests are administered at various points in the semesters that roughly coincide with points on the course where the skills to be tested have already been taught or practised. The course provides ample opportunity in the practice of these skills. Such an ordering on the learning continuum had implications for the content validity of the tests where,

> "Content validity refers to the ability of a test to measure what has been taught and subsequently learned by the students. It is obvious that teachers must see that the test is designed so that it contains items that correlate with the content of instruction. Thus it follows that unless students are given practice in oral communication in the foreign language classroom, evaluation of communication may not be valid...." Bartz (W H 1979).

By spacing out the tests in relation to the content, not only is the teacher-tester able to 'fit' the test to the content, she is also able after each test to obtain valuable feedback for the teaching of the subsequent domains that have been arranged in a cyclical fashion. Hence learning and performance is also on a cumulative basis because each skill taught and learnt or acquired presupposes and builds on the acquisition and the development of the preceding skills. It is on these bases that the tests have been developed and administered over a period of time. They are direct tests of performance that are communicative in nature and administered on a cumulative basis as part of on-going course assessment for both levels. The tests formats, and methods of elicitation owe much to some knowledge in the field (particularly the state of the art), test feedback, student introspection and teacher retrospection and experience with its full range of hunches and intuition.

Test Types

Level I

Level I as mentioned earlier consists of three test types.

1 Extended/'impromptu' speech

2 Group discussion

3 End-of-semester project

There are three speaking tasks of this type. Student speak for about 2 minutes on the first, 2-3 on the second and 3-5 on the third. The tasks test for three modes of speech as follows:

(i) Talking about oneself, others and experiences

(ii) Narrating and describing incidents and events

(iii) Expressing and justifying opinions.

1 (i) and (ii) are tested at the beginning of the first level mainly for diagnostic purposes as the students are of heterogeneous levels of proficiency. The speeches are staggered for both (i) and (iii) to ensure that each student has a minimum of a minute or so to prepare mentally for the topic. For (ii) they are all given an equal amount of time to prepare mentally and to make notes. When the testing begins they listen to each other speak, as the audience, thus providing the motivation and a 'valid' reason as it were for the task. (iii) is tested before the second half of the semester, to obtain information on learned behaviour as the students have had sufficient practice in expressing and justifying opinions through reaching consensus in group work. The topics for (i) and (ii) are well within the students' realm of experience and interest such as

The happiest day in my life.
The person who has influenced me the most.

However the topics for (iii) are of a slightly controversial nature such as

Should smoking be banned in all public places?
Do women make better teachers?

Both (ii) and (iii) are rated for global ability to communicate in the mode which is the overall ability of the student to persuade or justify reasons taken for a stand in the case of the latter and to describe, report and narrate in the case of the former.

2 The group discussion test is administered in the second half of the semester as by this time there has been plenty of practice in the interaction mode as the modus operandi of Level I is small group work. It tests specifically for oral interaction skills. The topics for group discussion tests are also based on the tacit principle that the content should be either familiar or known and not pose problems in the interaction process. Though the amount of communication (size of contribution) and substantiveness is rated as criteria, content per se is not rated. Group discussion in Level I tests lower order interaction skills that are discernible at the conversational level.

The groups discussion test has been modelled on the lines of the Bagrut group discussion test with some modifications (see Shomay, E., Reves, T. and Bejerano, Y. 1986 and Gefen, R. 1987). In Level I the topics are of matters that either concern or pose a problem to the test takers as UKM students. Hence there is sufficient impetus to talk about them and this 'guarantees' initiation by all members of the group in the discussion. Topics in the form of statements are distributed just before the tests from a prepared pool of topics. Each topic comes with a set of questions. Students are allowed to read the questions in advance but discussion on the topic and questions before the test is not permitted. These questions function as cues to direct and manage the interaction. They need not be answered. In fact students may want to speak on other aspects of the topic. An example of the topic and questions is as follows:

Scholarships should be awarded on need and not on merit.

(a) Are both equally important considerations?

(b) Should students have a say in who gets scholarships ie. have student representatives on scholarship boards?

(c) Do generous scholarships make students dependent on aid?

(d) Are repayable-upon-graduation loans better than scholarships as more students can benefit?

Groups are small and students are divided (depending on class size) into 4-5 (maximum) students per group. It has been possible to establish a rough ratio between rating time per test-taker and their number per group. Groups of 4

took 15-20 minutes to round off the discussion and groups of 5 took about 20-25 minutes. However, it is desirable not to cut off the discussion after 20-25 minutes, as extra time (usually an extra 5 minutes) helped to confirm ratings. Rating is immediate on the score sheets prepared for the test (see Appendix C ii). A variation of the topics with maximum backwash effect on learning is to use books that have been recommended for extensive reading as stimulus for group discussion. This has been trialled as a class activity.

It can be seen that the oral interview test is noticeably absent in the sampling of speech interactions for Level I of the course and probably begs the question why, as it is a common and well established test for testing oral interaction. Suffice to say that it is firstly one of the tests administered in the first year integrated program (and therefore sampled). Secondly the group discussion appears to be a more valid (face and content) test of oral interaction in relation to the course objectives.

3 Since a premium is placed on intelligibility/comprehensibility the end-of-semester project tests for overall verbal communicative ability in the rehearsed speech genre in the form of a news magazine that is audio taped for assessment and review. The news magazine may be presented either as a collage of items of news and views of events and activities on campus or thematically eg. sports on campus, cultural activities, student problems etc.

Level II

This level consists of 4 test types.

1 Group discussion

2 Public speaking

3 Debates

4 End-of-semester project

1 In the second level the group discussion test is administered early in the semester and the results used to determine how much more practice is needed in improving interaction skills before proceeding to the more formal performance-oriented speech genres. The topics for the group discussion in the second level are of a more controversial nature than in the first. Although cognitive load is expected to be greater in the tests, procedures for test administration and scoring are the same.

2 Public speaking is tested mid-way in the second semester after lecture-demonstrations and a series of class presentations. As a test of global communication skills, both verbal and non-verbal, it represents fairly high level order skills on the language learning continuum assumed for the course. Like debates, it is a sample of rehearsed speech in a formal situation. It is also viewed as a necessary advanced oral skill. Examples of topics are,

> Mothers should not go out to work.
> Alcoholism is a worse social evil than drug abuse.

3 The debate is placed at the end of the semester and usually viewed by the students as a finale of sorts of their oral communication skills. As with the public speaking test, topics and teams (for the debates) are made known well in advance and students work on the topics cooperatively for the latter. The backwash effect on the acquisition of social and study skills is tremendous as students are informed that ratings reflect group effort in the debating process. Both tests 2 and 3 are rated immediately and video taped for both review and record purposes.

4 The end-of-semester can take two forms --- that of a form of a platform activity (in the public speaking mode) or a chat show (speech interaction). Both test for skills learned or acquired during the course. The platform activity and the formal speech situation can be either an appeal (for blood donation, funds, etc) or the promotion of a product/service or idea. The chat show tests for oral interaction in the form of an extended interview of a 'celebrity'. Both tests simulate real life situations and allow for creativity and flexibility in that students can assume personae.

Criteria and Rating Scales

> "Testers should construct their own rating scales according to the purpose of the test". (Shohamy E. 1988)

Rating scales have been constructed for all the tests developed. A look at the criteria and the rating scales (see appendices) for the various tests discussed above, shows that the criteria for each test varies although some (mainly linguistic) recur as each test samples different types of communicative ability.

Working over a period of time (ie two years = four semesters) it has been possible to specify what criteria should be used to rate each test and therefore what sorts of rating scales to produce. It has also been possible to select specific

components from the broader criteria identified for each rating scale. In this sense each test has evolved pedagogically (mainly) and psychologically over a period of time to become more comprehensive in terms of the test (task) specifications. Feedback in the form of student responses (and reaction) to each task has also helped the tests to jell as they were used to make changes especially to the criteria and subsequently the rating scale so as to reflect a wider possible range of responses for each test.

Obviously comprehensiveness of criteria should not be at the expense of the feasibility of rating scales and the practicality of scoring procedures. Too many descriptors can make it difficult for a rater to evaluate the performance in any one task. Using all these simultaneously to make an immediate judgement is no mean task. Hence, instead of fully descriptive qualitative scales, more parsimonious rating scales were devised. Working hand in hand with a checklist of what are essentially holistic criteria which will vary according to test purpose, the tester rates analytically on a 1 to 4 or 6 point scale depending on the test. These scales are also grouped into 3 broad bands of 'weak', 'fair' and 'good' which provide guidelines to help the rater to keep on course in the absence of banded descriptors. There is also space on each score-sheet for tester comments. This allows the tester to make relevant remarks of each test on an individual basis particularly with reference to those factors that had an apparent effect on test performance, verbal, non-verbal or affective.

The problem (personal experience) with banded qualitative rating scales is that the descriptors may not fit the description of the individual student in that some of the performance variables for any one component may be absent while others may be present. And there are students whose performance defy 'pigeon-holing'. However, it is possible to categorize the same students, firstly, on a broad basis as 'weak', 'fair' and 'good' and then work from there to rate them analytically on weighted 6 point scales in this case. It may even be possible to describe them with reference to the criteria on an individual basis as it is small scale testing. While such rating procedures remain subjective and may even be criticized on that basis, at the very least they prevent stereo typing of students by not assigning their performance to prescriptive ready-made bands.

CONCLUSION

Test Anxiety

A certain amount of anxiety has been removed from the testing situations in the course firstly, because of the ongoing nature of the assessments and secondly because of the wider sampling of the speech genres.

> "There is ... evidence in the literature that the format of a task can unduly affect the performance of some candidates. This makes it necessary to include a variety of test formats for assessing each construct... In this case, candidates might be given a better chance of demonstrating potentially differing abilities (Weir, C. 1989).

Practitioners know that not only do levels of test anxiety vary from situation to situation and from testee to testee, it may not even be possible to eliminate anxiety as an affective variable. However, in order to further reduce test anxiety and to 'bias for best', students are informed at the beginning of each level about course objectives and expectations, test types and task specifications explained. Feedback is also provided after each test although actual scores obtained are not divulged.

Other Matters

All tests of courses on the university curriculum (cumulative or otherwise) are seen as achievement tests with scores and grades awarded accordingly. There is a certain amount of tension between rating according to specified criteria and the subsequent conversion of the weightage of the components of these criteria into scores. However despite this constraint it is still possible to speak of a student's profile of performance in the oral communication class from level to level. At the end of the second year similar judgements can be made of them as potential students for the B A in English Studies.

The oral communication course has also been offered more recently as an elective to other students and therefore involves more teachers. While the difference in clientele does change some of the course's methodological perspectives, the objectives have still been maintained as needs are broadly similar. The tests are now being subjected to a process of small-scale teacher validation since the question of some extrapolation is apparent. There have been informal training and practice sessions for the teachers in the use of the criteria and rating scales. Past samples of performance have been reviewed to arrive at bench marks and pre-marking sessions held to increase intra and inter-rater reliability. The intersubjectivity and teacher feedback on all these aspects are invaluable in improving the efficacy of the test as instruments, at least with reference to face and content validity. Obviously more work has to be done before anything conclusive can be said.

REFERENCES

Alderson, C. 1981. Report of the Discussion on Communicative Language Testing in Alderson, C. and Hughes, A. (eds) in ***Issues in Language Testing. ELT Documents III.*** *The British Council.*

Alderson, C. 1983. "Who Needs Jam?" ***Current Developments in Language Testing*** *London: Academic Press. 55-65.*

Bachman, L. (Forthcoming). ***Fundamental Considerations in Language Testing*** *Reading. Mass: Addison-Wesley.*

Bartz, W H. (1979). "Testing Oral Communication in the Foreign Language Classroom" in ***Language in Education: Theory and Practice.*** *Arlington. Virginia: Centre for Applied Linguistics 1-22.*

Brown, G. and Yule, G. 1983. ***Teaching the Spoken Language,*** *Cambridge: Cambridge University Press.*

Callaway, D R. 1980. "Accent and the Evaluation of ESL Oral Proficiency" in Wover, J. Perkins, K. ***Research in Language Testing,*** *Rowley. Mass: Newbury House.*

Carroll, B J. 1980. ***Testing Communicative Performance,*** *Oxford: Pergamon Press.*

Clark, J L D. 1975. "Theoretical and Technical Considerations in Oral Proficiency Testing in Jone, R. and Spolsky, B. (eds) ***Testing Language Proficiency*** *Arlington. Virginia: Centre for Applied Linguistic 10-28.*

Gefen, R. 1987. "Oral Testing --- The Battery Approach" ***Forum*** *25.2 24-27*

Kreshen, S. and Terrell T. 1983. ***The Natural Approach: Language Acquisition in the Classroom,*** *Pergamon: Oxford.*

Nair-Venugopal, S. 1988. "Simulations, authentic oral language use in multiracial settings" ***Babel Occasional Papers*** *1 27-31*

Skehan, P. 1988. Language Testing: Part I 1-10. Language Testing: Part II 1-13. ***Language Teaching Abstracts.***

Shohamy, E. 1988. "A Proposed Framework for Testing the Oral Language of Second/Foreign Language Learners" ***Studies in Second Language Acquisition*** *10.2 165-180*

Shohamy, E., Reves. T. and Bejerano, Y. 1986. "Large Scale Communicative Language Testing: A Case Study" in Lee, Fok, Lord and Low (eds) ***New Directions in Language Testing*** *Oxford: Pergamon Press.*

Spolsky, B. 1975. "Language Testing: Art or Science" Lecture Delivered ***AILA World Congress '75*** *University of Stuttgart.*

Weir, C. 1988. "Approaches to Language Test Design: A Critical Review" Lecture Delivered at British Council Specialist Course 937 ***Testing Oral Interaction: Principles and Practice,*** *Reading, England.*